Science is broken

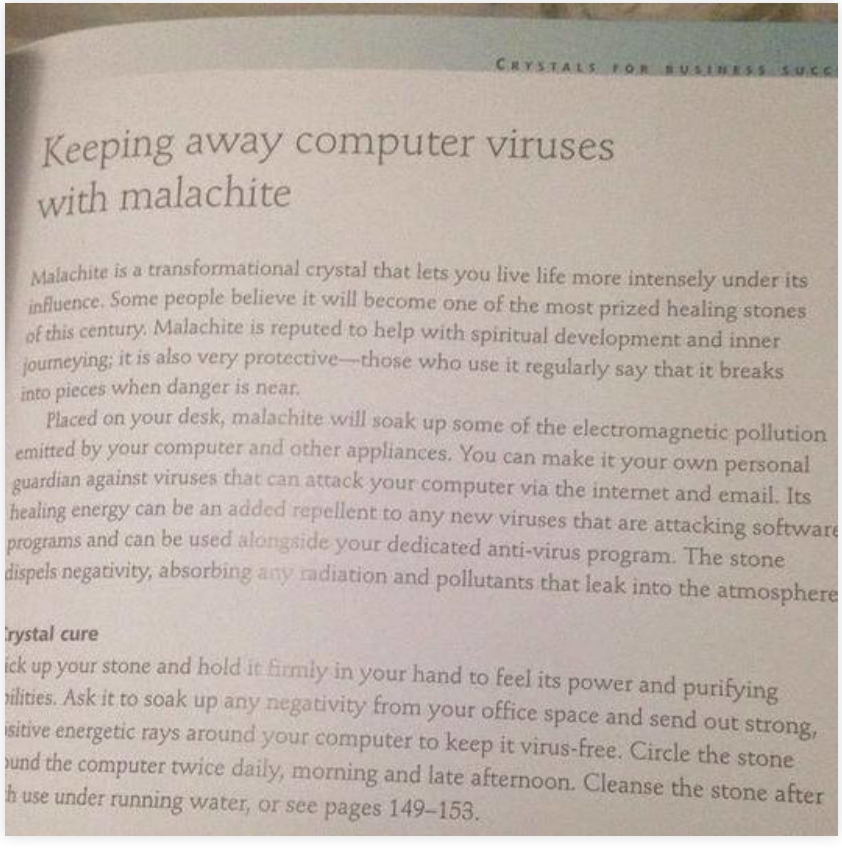
Hanno Böck

https://betterscience.org/

https://hboeck.de/

Can we trust the scientific method?

A simple example



CRYSTALS FOR BUSINESS SUCC

Keeping away computer viruses with malachite

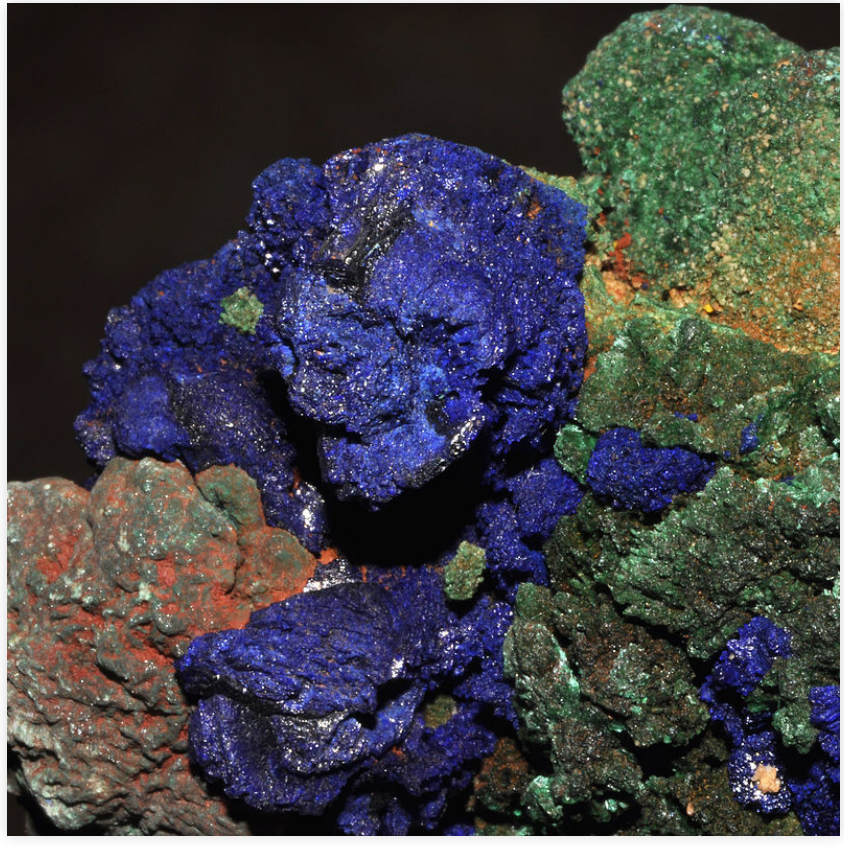
Malachite is a transformational crystal that lets you live life more intensely under its influence. Some people believe it will become one of the most prized healing stones of this century. Malachite is reputed to help with spiritual development and inner journeying; it is also very protective—those who use it regularly say that it breaks into pieces when danger is near.

Placed on your desk, malachite will soak up some of the electromagnetic pollution emitted by your computer and other appliances. You can make it your own personal guardian against viruses that can attack your computer via the internet and email. Its healing energy can be an added repellent to any new viruses that are attacking software programs and can be used alongside your dedicated anti-virus program. The stone dispels negativity, absorbing any radiation and pollutants that leak into the atmosphere

rystal cure

ick up your stone and hold it firmly in your hand to feel its power and purifying ilities. Ask it to soak up any negativity from your office space and send out strong, sitive energetic rays around your computer to keep it virus-free. Circle the stone und the computer twice daily, morning and late afternoon. Cleanse the stone after h use under running water, or see pages 149–153.

Reddit



Parent Géry, Wikimedia Commons

Let's do a study!

We'll do a randomized controlled trial (RCT), which is the gold standard in many fields of science.

Do Malachite crystals prevent malware infections?

Study design (RCT, part 1)

- Take a group of 20 computer users.
- Split them randomly in two groups.

Study design (RCT, part 2)

- Give one group a malachite crystal to put on their desk.
- Give the other group a fake malachite crystal that cannot be easily distinguished from a real one (control group).
- After 6 months check how many malware infections they had.

Simulate study with random data

```
#!/usr/bin/env python3

import os
import numpy
from scipy import stats

a = [float(os.urandom(1)[0] % 4) for _ in range(10)]
b = [float(os.urandom(1)[0] % 4) for _ in range(10)]

print("%s\n%s" % (a, b))

t, p = stats.ttest_ind(a, b)

print("%.2f;%.2f;%.2f" % (numpy.mean(a), numpy.mean(b), p))
```

p-value

A p-value is the probability that you get a false positive result in idealized conditions if there is no real effect.

In many fields of science $p<0.05$ is considered significant.

Malachite	Fake	p-value
1.40	1.50	0.87
2.10	1.70	0.40
1.50	1.10	0.44
2.10	1.30	0.12
1.10	1.90	0.11
1.20	1.20	1.00
1.80	2.40	0.12
1.70	2.00	0.58
1.20	1.70	0.30
2.10	1.20	0.06

Malachite	Fake	p-value
1.60	1.60	1.00
1.80	1.80	1.00
1.30	1.50	0.72
1.70	1.10	0.25
1.40	1.70	0.49
1.70	1.60	0.83
1.80	0.80	0.03
1.60	1.30	0.61
0.80	1.30	0.30
1.00	1.60	0.28

Malachite	Fake	p-value
1.40	1.50	0.87
2.10	1.70	0.40
1.50	1.10	0.44
2.10	1.30	0.12
1.10	1.90	0.11
1.20	1.20	1.00
1.80	2.40	0.12
1.70	2.00	0.58
1.20	1.70	0.30
2.10	1.20	0.06

Malachite	Fake	p-value
1.60	1.60	1.00
1.80	1.80	1.00
1.30	1.50	0.72
1.70	1.10	0.25
1.40	1.70	0.49
1.70	1.60	0.83
1.80	0.80	0.03
1.60	1.30	0.61
0.80	1.30	0.30
1.00	1.60	0.28

The effects of Malachite crystals on Malware Infections

Hanno Böck

December 20, 2017

Abstract

We performed a randomized controlled trial about the effects of malachite crystals on the number of malware infections. Participants of our study that put a malachite crystal on their desk were affected by significantly more malware infections ($p = 0.03$) than participants in a control group.

We just created a significant result out of random data

Publication Bias

What is stopping scientists from doing this?

Usually nothing!

Let's look at a real example: SSRIs (Antidepressants)

Publication Bias and Antidepressants

- 74 studies on SSRIs, data from the FDA.
- 37 out of 38 studies with positive results published.
- 14 out of 36 studies with negative results published, of those 11 claimed a positive outcome.

Turner et al. 2008, NEJM

With Publication Bias you can create results out of nothing.

But it's not efficient, you need 20 studies on average to get a result.

How to interpret our results?

In a scientific study many decisions have to be made:

- What to do with dropouts?
- What to do with cornercase results?
- What exact outcome are we looking for?
- What variables do we control for?

Each of these decisions has a small impact on the result

p-Hacking

Even if there is no real result one of these variations may cause enough skew to be significant.

This may be a subconscious process

- Scientists don't start and say: "Today I'm gonna p-hack my result."
- They may subconsciously favor decisions that look like they may lead to the result they expect.

What stops scientists from p-Hacking?

Usually nothing.

Conclusion

The scientific method is a way to create evidence for whatever theory you like.

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on

Ioannidis, PLOS Medicine, 2005

Flávio Britto Calil, Wikimedia Commons

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

A lot of things were wrong with this study.

But it was absolutely in line with the existing standards in experimental psychology.

Francis 2012, Psychonomic Bulletin & Review

Psychology is facing a Replication Crisis

Many effects of psychology that were considered facts failed to replicate.

Slate

GREEN ROOM

NEWS AND COMMENTARY ABOUT ENVIRONMENTAL ISSUES.

DEC. 7 2009 6:24 PM

Buy Local, Act Evil

Can organic produce and natural shampoo turn you into a heartless jerk?

By Rebecca Tuhus-Dubrow

As the owner of several energy-efficient light bulbs and a recycled umbrella, I'm familiar

become a supporter

theguardian

news | opinion | sport | arts | lifestyle

environment ▸ climate change | wildlife

How going green may make you mean

Ethical consumers less likely to be kind and more likely to steal, study finds

- You ask, they answer: Ethical Consumer magazine
- Julian Baggini: Goodies behaving badly

Kate Connolly in Berlin

Mon 15 Mar '10 19.42 GMT

486 255

This article is 7 years old

When Al Gore was caught running up huge energy bills at home at the same time as lecturing on the need to save electricity, it turns out that he was

VANCOUVER SUN

Study shows 'Green' shoppers more likely to cheat

MIKE BARBER, CANWEST NEWS SERVICE 12.19.2009 |

Man shops for organic produce at a Whole Foods Market in San Francisco. *JUSTIN SULLIVAN / GETTY IMAGES*

If buying an organic apple instead of one caked in pesticides eases your conscience, there's a good chance that your next ethical decision might not be a good one.

According to the results of a University of Toronto study, participants who assigned more social value to 'green' shopping were more likely to cheat

RELATED

Try uncorking organic wines for your holiday parties

Looking good in green: The basics of sustainable fashion

Urban, Bahník, Kohlová 2017, PsyArXiv

A warning

Don't be too snarky about psychologists. Your field is probably not any better. You just don't know yet.

Other fields have a replication crisis as well

Pharma company Amgen failed to replicate 47 out of 53 preclinical cancer studies in 2012.

(Though there are a few problems with this result.)

Some fields don't have a replication problem - because nobody is trying to replicate results.

What can be done about all this?

The scientific process from analysis to publication needs to be decoupled from its results.

Preregistration

Preregistration

Announce in a public registry what you plan to do in your research.

Later people can check if you published your results and if you changed your research on the way.

This is typically done in drug trials.

It doesn't work very well - but it's better than nothing.


AllTrials


COMPARE
TRACKING SWITCHED OUTCOMES IN CLINICAL TRIALS

We know Big Pharma is bad

But think about this: Whenever you read about problems in drug trials you should consider that most other fields don't do preregistration at all.

Right now there's a trend that people from computer science want to change medicine (Big Data / ML).

Some people in medicine are very worried about this - because the computer science people bring their weak scientific standards with them.

Registered Reports

Open letter in The Guardian, 2013

Registered Reports

Turn scientific publication process upside down.

- First publish a protocol for your experiment to a scientific journal.
- Journal decides on publication based on the protocol before the results are in.
- Publish results - independent of outcome.

Other improvements

- Sharing of data, code, methods.
- Large-scale collaboration (one well-designed large study is better than many small ones).
- Higher statistical threshold ($p<0.05$ means practically nothing).

How's my field doing?

- Are statistical results preregistered in any way?
- Are negative results usually published?
- Are there independent replications of all relevant results?

If you answer all these questions with "No" you are probably not doing science.

You're the alchemists of our time.

Bad incentives

- Citation counts (Impact Factor).
- Publicity.

Existing incentives strongly favor interesting results - not correct results

Isn't science self-correcting?

If you confront scientists with evidence for Publication Bias and p-hacking - surely they'll immediately change their practices. That's what scientists do, right?

There is some evidence that in fields where statistical tests of significance are commonly used, research which yields nonsignificant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs—an "error of the first kind"—and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

1959

Theodore Sterling, American Statistical Association

This article presents evidence that published results of scientific investigations are not a representative sample of results of all scientific studies. [...] These results also indicate that practice leading to publication bias have not changed over a period of 30 years.

Sterling 1995, The American Statistician

If science is self-correcting it's pretty damn slow in doing so.

Are you prepared for boring science?

There is a choice between TED-talk science and boring science.

TED-talk science

- Mostly positive and surprising results.
- Large effects.
- Many citations.
- Media attention.
- You may be able to give a TED talk about it.
- Usually not true.

Boring science

- Mostly negative results.
- Small effects.
- Boring.
- Closer to the truth.

I prefer boring science.

But this is a tough sell.

Thanks for listening!

https://betterscience.org/
https://hboeck.de/